Reinforcement Learning Possibilita Maior Retorno que Otimização Estática? Uma Análise para Criptomoedas no Período de 2021 a 2024

Does Reinforcement Learning Offer a Higher Return Than Static Optimization? An Analysis for Cryptocurrencies in the Period 2021-2024

Vitor Augusto Ferreira [1]

Helberte João França Almeida [2]

Adilson Giovanini [3]

Resumo: O presente trabalho busca utilizar técnicas de *Machine Learning*, especificamente o *Reinforcement Learning* (Aprendizado por Reforço), para otimizar a alocação de recursos em um portfólio composto por cinco criptomoedas de grande capitalização: *Bitcoin* (BTC), *Ethereum* (ETH), *Binance Coin* (BNB), *Ripple* (XRP) e Solana (SOL) no período de 2021 a 2024. A aplicação do teste não paramétrico de *Kruskal-Wallis* revelou a ausência de diferença estatisticamente significante (p-valor > 0,05) entre as abordagens em todos os cenários. Os resultados obtidos sinalizam que o modelo de *Reinforcement Learning* superou as estratégias tradicionais no *Bull Market*, com um retorno acumulado de 73,26%. Contudo, o modelo enfrentou dificuldades no *Bear Market* e no *Sideways Market*, o qual não conseguiu superar a alocação estática. Portanto, conclui-se que o modelo é eficaz em mercados de alta, mas requer ajustes para melhorar seu desempenho em cenários adversos.

Palavras-chave: Otimização de portfólio; Criptoativos; Machine Learning; Volatilidade.

Abstract: *This work seeks to use Machine Learning techniques, specifically Reinforcement Learning, to optimize resource allocation in a portfolio composed of five large-cap cryptocurrencies: Bitcoin (BTC), Ethereum (ETH), Binance Coin (BNB), Ripple (XRP), and Solana (SOL) in the period of 2021 the 2024. The results indicate that the Reinforcement Learning model outperformed traditional strategies in the bull market, with a cumulative return of 73.26%, however, struggled in the bear market and sideways market, failing to outperform static allocation. Therefore, conclusion is that the model is effective in bull markets but requires adjustments to improve its performance in adverse scenarios.*

Keywords: *Portfolio Optimization; Cryptoassets; Machine Learning; Volatility.*

Classificação JEL: C60; C61; C63; G11.

[1] Graduado em Economia, Universidade Federal de Santa Catarina (UFSC).

[2] Programa de Pós Graduação em Economia - PPGECO, Universidade Federal de Santa Catarina (UFSC).

[3] Universidade do Estado de Santa Catarina (UDESC).

1. Introdução

O mercado de criptoativos, impulsionado pelo surgimento do *Bitcoin* em 2009 (Nakamoto, 2008), consolidou-se como uma das inovações financeiras mais significativas do século XXI. Desde então, o ecossistema de criptomoedas expandiu-se exponencialmente, atingindo uma capitalização de mercado global que superou US$ 1 trilhão em 2021. Essa expansão foi acompanhada pela diversificação de ativos digitais, como *Ethereum* (Buterin, 2014), *Binance Coin*, *Ripple* e *Solana*, cada um com funcionalidades únicas, como contratos inteligentes, *tokens* não fungíveis (*NFTs*) e soluções de escalabilidade. Contudo, apesar do crescimento e da adoção crescente por parte de instituições financeiras e governos, o mercado de criptoativos permanece marcado por alta volatilidade, falta de regulação uniforme e sensibilidade a eventos macroeconômicos e geopolíticos (SChär, 2021).

A conjuntura global tem influenciado o comportamento dos criptoativos. Eventos como crises financeiras, mudanças nas políticas monetárias de grandes economias (como os Estados Unidos e a União Europeia), tensões geopolíticas e até mesmo declarações de figuras públicas têm impactado diretamente os preços desses ativos. Assim, o aumento nas taxas de juros pelo *Federal Reserve* em 2022 e 2023 levou a uma fuga de capitais de ativos de risco, incluindo criptomoedas, resultando em um prolongado *Bear Market*. Por outro lado, a adoção de criptomoedas por países como El Salvador como moeda legal e o interesse de grandes corporações em *blockchain* e *Web3* têm impulsionado ciclos de alta, conhecidos como *Bull Markets*. Essa dualidade entre momentos de euforia e desconfiança torna o mercado de criptoativos particularmente complexo e desafiador para investidores.

A otimização de portfólios, uma prática consolidada em mercados tradicionais, ganha relevância no universo dos criptoativos. No entanto, métodos clássicos, como a Teoria Moderna do Portfólio de Markowitz, que busca maximizar o retorno para um determinado nível de risco, mostram-se limitados diante da dinâmica única desse mercado (Markowitz, 1952). A alta volatilidade, a correlação não linear entre ativos e a influência de fatores externos exigem abordagens mais sofisticadas e adaptativas. Nesse cenário, técnicas de *Machine Learning*, em particular o *Reinforcement Learning*, emergem como ferramentas promissoras (Sutton; Barto, 1997). Essas técnicas permitem a criação de modelos que aprendem com os dados, ajustando estratégias de alocação em tempo real e adaptando-se às mudanças de mercado de forma dinâmica.

Além disso, a integração de *Machine Learning* na gestão de portfólios de criptoativos está alinhada com a tendência global de uso de inteligência artificial e análise de dados em finanças. Plataformas como *hedge funds* e gestoras de recursos têm adotado algoritmos de *ML* para prever movimentos de mercado, identificar padrões e otimizar decisões de investimento. Cabe ressaltar que as aplicações das técnicas de *ML* ao mercado de criptomoedas ainda são incipientes, especialmente no que diz respeito à otimização de portfólios. Esse *gap* na literatura e na prática reforça a relevância deste trabalho, que busca explorar o potencial do *Reinforcement Learning* para superar as limitações das abordagens tradicionais.

O presente estudo possui o seguinte problema de pesquisa: Em que medida técnicas de *Reinforcement Learning* conseguem otimizar o retorno ajustado ao risco em portfólios de criptoativos, em comparação com estratégias tradicionais, diante da alta volatilidade e dos diferentes cenários de mercado (*Bull*, *Bear* e *Sideways*)? Esse problema surge da necessidade de superar as limitações das abordagens tradicionais de otimização de portfólio, que apresentam limitações para lidar com a volatilidade e a imprevisibilidade dos criptoativos.

O artigo tem como objetivo utilizar técnicas de *Machine Learning*, especificamente o *Reinforcement Learning*, para otimizar a alocação de recursos em um portfólio

composto por cinco criptomoedas de grande capitalização: *Bitcoin*, *Ethereum*, *Binance Coin*, *Ripple* e *Solana* no período de 2021 a 2024. O motivo pela escolha destas criptomoedas é devido ao fato que estas representam 95% das transações neste no mercado de cripoativos. Para alcançar o objetivo de pesquisa, utiliza-se da abordagem de *Machine Learning*, com foco na aplicação de *Reinforcement Learning* para maximizar o retorno ajustado ao risco. Precisamente, o estudo contribui para a prática de investimentos em criptoativos, fornecendo uma ferramenta quantitativa e adaptativa que pode auxiliar investidores a tomar decisões mais eficientes. A disponibilização do código-fonte e dos dados utilizados no *Google Colab* também promove a replicabilidade e a transparência do estudo, incentivando futuras pesquisas na área.

Além desta introdução, o estudo encontra-se estruturado em mais quatro seções. A segunda seção apresenta o referencial teórico e estudos correlatos ao tema abordado. Por sua vez, a terceira seção traz a metodologia, bem como a implementação computacional. Em seguida, a quarta seção mostra os resultados obtidos. Por fim, a quinta seção apresenta as considerações finais.

2. Referencial teórico

2.1. Otimização de Portfólio

A otimização de portfólio é um processo central na gestão de investimentos, cujo objetivo é selecionar e balancear ativos financeiros de forma a maximizar os retornos esperados enquanto se minimiza o risco associado. Esse conceito foi formalizado por Harry Markowitz em 1952, com a Teoria Moderna do Portfólio (TMP), que apresentou uma maneira como investidores e gestores abordam a alocação de ativos (Markowitz, 1952).

A TMP parte do princípio de que o risco de um portfólio não depende apenas do risco individual de cada ativo, mas também da forma como esses ativos se relacionam entre si. Essa relação é medida pela covariância, que indica o grau em que os retornos de dois ativos se movem em conjunto. Assim, ao combinar ativos que não estão perfeitamente correlacionados é possível reduzir o risco total do portfólio sem penalizar os retornos, um conceito conhecido como diversificação. A diversificação é a base da construção de portfólios eficientes, que buscam equilibrar risco e retorno de forma ótima.

Um dos pilares da TMP é a fronteira eficiente, que representa o conjunto de portfólios que oferecem o maior retorno esperado para um determinado nível de risco. Portfólios que situam-se abaixo dessa fronteira são considerados subótimos, pois não conseguem alcançar o melhor equilíbrio possível entre risco e retorno. Para construir a fronteira eficiente, é necessário calcular a variância e a covariância dos ativos, o que permite determinar o risco total do portfólio. A variância mede a volatilidade individual de cada ativo, enquanto a covariância captura a interdependência entre os movimentos de preços de dois ativos. A matriz de covariância é um elemento essencial no processo de otimização, pois permite quantificar o impacto da diversificação no risco do portfólio (Markowitz, 1952).

O Modelo *CAPM* (*Capital Asset Pricing Model*) complementa a TMP ao introduzir o conceito de risco sistemático e não sistemático. O risco sistemático conhecido como o risco de mercado, é inerente a todo o mercado e não pode ser eliminado através da diversificação. Já o risco não sistemático é específico de um ativo ou setor e pode ser reduzido com a combinação adequada de ativos. O *CAPM* utiliza o beta como medida de sensibilidade de um ativo em relação ao mercado. Desta forma, um beta

maior que um, indica que o ativo tende a ser mais volátil que o mercado. Por sua vez, um beta menor que um, sugere menor volatilidade. O *CAPM* é amplamente utilizado para estimar o retorno esperado de um ativo com base em seu risco sistemático, mas enfrenta críticas, especialmente em mercados altamente voláteis ou com correlações instáveis, como o de criptoativos (Fama; French, 1992).

A otimização de portfólio muitas vezes envolve a consideração de restrições práticas, como liquidez, fatores setoriais ou limitações legais. Em mercados com baixa liquidez, a compra ou venda de grandes volumes de ativos pode impactar significativamente seus preços, o que deve ser considerado no processo de otimização. Além disso, restrições legais e regulatórias podem limitar a exposição a certos ativos ou setores. Para lidar com essas restrições, algoritmos de otimização são frequentemente empregados. Um dos métodos mais comuns é o gradiente descendente, que busca encontrar o mínimo de uma função de risco sujeita a determinadas restrições. Esses algoritmos permitem a solução de problemas complexos de otimização, especialmente em portfólios com muitos ativos, no qual a solução manual se torna inviável.

Apesar da robustez da TMP e do *CAPM*, esses modelos tradicionais enfrentam críticas e desafios significativos. Uma das principais críticas é a dependência excessiva de dados históricos para estimar variâncias, covariâncias e retornos esperados. Em mercados altamente dinâmicos, como o de criptoativos, as correlações entre ativos podem mudar rapidamente, tornando as estimativas baseadas em dados históricos pouco confiáveis. Além disso, a TMP assume que os retornos dos ativos seguem uma distribuição normal, o que nem sempre é verdadeiro, especialmente em mercados com eventos extremos e caudas longas, como os criptoativos (Schär, 2021). Essas limitações têm impulsionado o desenvolvimento de novas abordagens, como o modelo *Black-Litterman*, que incorpora visões subjetivas do investidor ao processo de otimização, e o *Risk Parity*, que busca balancear o portfólio com base no risco contribuído por cada ativo, em vez de focar apenas no retorno esperado (Black; Litterman, 1992).

2.2. Criptoativos

Os criptoativos emergiram como uma resposta à crise financeira de 2008, marcando o início de um novo sistema monetário e tecnológico. O *Bitcoin* lançado em 2009 por uma entidade ou grupo sob o pseudônimo de Satoshi Nakamoto, foi a primeira criptomoeda a resolver o problema do gasto duplo sem a necessidade de uma autoridade central. Essa inovação foi possível graças à tecnologia *blockchain*, um livro-razão distribuído que registra transações de forma segura, transparente e imutável. Assim, o *BTC* foi concebido como uma alternativa ao sistema financeiro tradicional, oferecendo uma moeda digital descentralizada.

O sucesso do *Bitcoin* inspirou o surgimento de outras criptomoedas, mas foi o *Ethereum*, lançado em 2015 por Vitalik Buterin que trouxe uma inovação revolucionária: os contratos inteligentes. Diferente do *BTC*, que foi projetado principalmente como uma moeda digital, o *Ethereum* foi concebido como uma plataforma programável, capaz de executar aplicações descentralizadas (*dApps*) por meio de contratos inteligentes. Esses contratos são códigos autônomos que executam automaticamente acordos pré-definidos quando certas condições são atendidas, eliminando a necessidade de intermediários. Essa funcionalidade abriu caminho para uma infinidade de aplicações, desde finanças descentralizadas (*DeFi*) até *tokens* não fungíveis (*NFTs*), transformando o *ETH* em uma das *blockchains* mais importantes do ecossistema cripto (buterin, 2014).

A tecnologia *blockchain*, que sustenta tanto o *BTC* quanto o *ETH*, é um dos pilares fundamentais dessa revolução. Precisamente, esta tecnologia trata-se de uma rede distribuída de nós que validam e registram transações em blocos encadeados cronologicamente. Cada bloco contém um conjunto de transações, e uma vez adicionado à cadeia, não pode ser alterado sem o consenso da rede. Essa imutabilidade garante segurança e transparência, enquanto a descentralização elimina pontos únicos de falha e reduz a dependência de intermediários. No caso do *Bitcoin*, a *blockchain* é mantida por um mecanismo de consenso chamado *Proof of Work* (*PoW*), no qual mineradores competem para resolver problemas matemáticos complexos e validar transações.

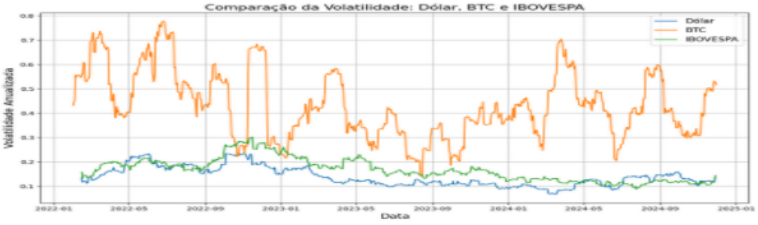
Por sua vez, o *Ethereum*, inicialmente baseado no *PoW*, migrou para o *Proof of Stake* (*PoS*) em 2022, visando reduzir o consumo de energia e aumentar a escalabilidade. Essa transição foi um marco importante para a sustentabilidade do ecossistema cripto, uma vez que o *PoW* do *Bitcoin* tem sido criticado por seu alto consumo energético, comparável ao de pequenos países. A busca por soluções mais sustentáveis, como o *PoS* e outras alternativas de consenso, tem sido uma prioridade para o setor, especialmente diante das crescentes preocupações ambientais (Zheng *et al*., 2017).

Os *NFTs* (*Non-Fungible Tokens*) são outro desenvolvimento significativo no ecossistema cripto, impulsionado principalmente pelo *Ethereum*. Diferente das criptomoedas tradicionais, que são fungíveis (ou seja, cada unidade é intercambiável), os *NFTs* são únicos e indivisíveis, representando propriedade sobre itens digitais ou físicos. Essa tecnologia permitiu a tokenização de arte digital, colecionáveis, itens de jogos e até mesmo imóveis, criando um novo mercado para bens digitais. O *boom* dos *NFTs* em 2021, com vendas recordes de obras de arte digitais e colecionáveis, destacou o potencial dessa tecnologia para transformar indústrias como entretenimento, arte e *gaming*. No entanto, também levantou questões sobre sustentabilidade, devido ao alto consumo de energia e sobre a especulação excessiva no mercado. A migração do *Ethereum* para o *PoS* ajudou a mitigar parte dessas preocupações, mas o debate sobre o impacto ambiental dos *NFTs* e dos criptoativos em geral continua relevante (Dowling, 2022).

Além do *BTC* e do *ETH*, outros criptoativos como *Solana*, *Binance Coin* e *XRP* também desempenham papéis importantes no ecossistema. O *Solana*, por exemplo, ganhou destaque por sua alta velocidade e baixas taxas de transação, utilizando um mecanismo híbrido de consenso que combina *Proof of History* (*PoH*) com *Proof of Stake* (*PoS*). A criptomoeda *BNB*, criado pela *exchange Binance*, evoluiu de um *token* utilitário para uma criptomoeda multifuncional, sendo utilizado na *Binance Smart Chain* (*BSC*), uma *blockchain* compatível com contratos inteligentes. O *XRP* foi desenvolvido pela *Ripple Labs* com foco em transações internacionais, oferecendo uma alternativa rápida e eficiente ao sistema *SWIFT*. Cada um desses criptoativos trouxe inovações específicas, mas também enfrentou desafios, como questões de escalabilidade, centralização e regulamentação (Schär, 2021).

A volatilidade é uma característica marcante dos criptoativos, influenciada por fatores como especulação, adoção institucional, mudanças regulatórias e eventos macroeconômicos. A título de exemplificação, o *BTC* já passou por ciclos de alta extrema seguida por correções significativas, refletindo a natureza incipiente e altamente sensível desse mercado. A volatilidade pode ser vista como uma oportunidade para investidores em busca de retornos expressivos, mas também como um risco para aqueles que buscam estabilidade. À medida que o mercado amadurece e a utilização aumenta, espera-se que a volatilidade diminua, embora ainda seja um aspecto central a ser considerado por qualquer participante do ecossistema cripto (Bohme *et al*., 2015).

Figura 01 – Volatilidade - Dólar, BTC e IBOV



Fonte: Elaboração própria.

2.3. *Machine Learning*

O aprendizado de máquina é uma área da inteligência artificial que tem suas raízes em décadas de pesquisa e desenvolvimento, com marcos históricos que remontam à década de 1950, quando os primeiros conceitos de redes neurais e algoritmos de aprendizado começaram a surgir. A historicidade da tecnologia é marcada por períodos de avanços e retrocessos, conhecidos como 'invernos da IA', no qual o entusiasmo inicial era seguido por desilusões devido às limitações computacionais e teóricas da época. No entanto, a partir dos anos 2000, com o advento de melhores *hardwares*, grandes volumes de dados e avanços em técnicas de aprendizado profundo (*Deep Learning*), o campo experimentou um renascimento.

Esse período, muitas vezes chamado de '*boom* da IA', trouxe consigo a popularização de técnicas como redes neurais convolucionais (*CNNs*) para visão computacional, redes neurais recorrentes (*RNNs*) para processamento de linguagem natural e, mais recentemente, transformadores (*Transformers*) que revolucionaram a área. Dentro desse contexto, o *Reinforcement Learning* emergiu como uma abordagem para problemas de tomada de decisão sequencial, ganhando destaque com vitórias em jogos complexos como *Go* e xadrez, além de aplicações em robótica, finanças e controle de sistemas autônomos. (Goodfellow *et al*., 2016).

Os tipos de aprendizado de máquina podem ser categorizados em três grandes grupos: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço. No aprendizado supervisionado, o modelo é treinado com um conjunto de dados rotulados, no qual cada exemplo de entrada está associado a uma saída desejada. Esse tipo de aprendizado é amplamente utilizado em tarefas como classificação e regressão. Por sua vez, o aprendizado não supervisionado lida com dados não rotulados, buscando identificar padrões ou estruturas intrínsecas, como em técnicas de *clustering* ou redução de dimensionalidade.

O *Reinforcement Learning* difere dessas abordagens por não depender de um conjunto de dados estático. Assim, o agente aprende interagindo com um ambiente dinâmico, recebendo *feedback* na forma de recompensas ou penalidades, e ajustando sua política de decisão para maximizar a recompensa cumulativa ao longo do tempo. Essa característica faz do *RL* uma ferramenta poderosa para problemas em que a tomada de decisão sequencial é essencial, como em jogos, controle de robôs e otimização de processos (Sutton; Barto, 2018).

O funcionamento do *Reinforcement Learning* é baseado na interação contínua entre um agente e um ambiente, modelado como um Processo de Decisão de Markov (*MDP*). Em um *MDP*, o ambiente é descrito por um conjunto de estados, ações, recompensas e uma função de transição que define como o estado evolui em resposta às ações do agente. O objetivo do agente é aprender um comportamento, uma função que mapeia estados para ações, que maximize a recompensa esperada ao longo do

tempo. Para isso, o agente utiliza técnicas como funções de valor, que estimam o retorno esperado a partir de um determinado estado ou par estado-ação, e métodos de política, que buscam otimizar diretamente a política de decisão.

Um conceito fundamental no *RL* é o de exploração *versus* explotação, no qual o agente deve equilibrar a exploração de novas ações para descobrir estratégias mais eficientes com a exploração de ações conhecidas que já produzem recompensas satisfatórias. Esse equilíbrio é crucial para o sucesso do aprendizado, especialmente em ambientes complexos e estocásticos (Sutton; Barto, 2018).

A literatura recente sobre otimização de portfólios de criptoativos via *Reinforcement Learning* tem enfatizado funções-objetivo ajustadas ao risco (p.ex., *Sharpe*, *CVaR*, *drawdown*), dada a presença de caudas pesadas e regimes de alta volatilidade nesse mercado. Um avanço central é a incorporação direta de medidas de risco à recompensa do agente, substituindo o alvo clássico de retorno puro por métricas que refletem o *trade-off* risco-retorno.

No escopo estritamente relacionado à otimização de carteiras de criptomoedas, Cui *et al.* (2023), motivado pelas caudas longas acentuadas das criptomoedas, propõem uma abordagem *CVaR-based deep reinforcement learning* para construção de portfólios. Desta forma, utilizando dados de 2015–2021, os autores mostram que o arcabouço baseado em *CVaR* supera técnicas tradicionais, tanto na formação de fronteiras eficientes quanto em métricas de desempenho ajustadas ao risco, destacando que o *framing* com *CVaR* auxilia o agente a mitigar perdas extremas sem abdicar de rentabilidade.

Em linha com a priorização de recompensas ajustadas ao risco, estudos metodológicos em *DRL* de portfólio para ativos fora do universo das criptomoedas reforçam o uso do *Sharpe* como recompensa incremental, viabilizando o aprendizado por passo de tempo e aproximando o *Sharpe* total ao final do horizonte. Neste contexto, Sood *et al.* (2023) estruturam e comparam um *policy-gradient*, demonstrando ganhos em *Sharpe* e retorno absoluto quando a recompensa maximiza diretamente o retorno ajustado ao risco. Esse desenho de recompensa tem sido reutilizado ou estendido em propostas recentes, com mecanismos de auto-recompensa com *Sharpe* de curto prazo (k-dias) para estabilizar aprendizado e capturar risco local (Huang *et al*., 2024).

Há também evidências de *frameworks risk-aware* generalizáveis que combinam múltiplas funções de recompensa (*log-returns*, *differential Sharpe*, *MDD*) para produzir uma política unificada mais estável em regimes não estacionários. Embora não restritos a cripto, Mishra, Srivastava e Singh (2025) reportam melhorias de desempenho ajustado ao risco e são relevantes para replicação no contexto de criptoativos. Por fim, revisões e levantamentos recentes reforçam que, no universo das criptomoedas, a avaliação por métricas ajustadas ao risco (*Sharpe*, *CVaR*, *MDD*) tornou-se prática dominante, tanto em *trading* quanto em alocação, justamente pela maior incidência de assimetrias e eventos de cauda (Liu *et al*., 2023; Jiang; Liang, 2017).

A aplicação de técnicas de aprendizado por reforço tem ampliado a eficiência de modelos de decisão sob restrições de custo. Neste contexto, Gu, Y. *et al.* (2024) propõem um modelo de escalonamento de *workflows* em nuvem baseado em *DRL*, incorporando sensibilidade a custos no processo de otimização. Por sua vez, Liu, P. (2024) demonstra que a otimização bayesiana *cost-aware* melhora o rebalanceamento de portfólios ao reduzir custos de transação e elevar a eficiência alocativa, evidenciando a centralidade das fricções nos problemas de decisão dinâmica.

A incorporação de mecanismos adaptativos torna-se fundamental em ambientes financeiros sujeitos a instabilidades estruturais. Lu, J. Y. *et al.* (2022) desenvolvem uma estratégia de *pairs trading* sensível a quebras estruturais, ampliando a robustez do modelo em períodos de elevada volatilidade. De modo complementar, Conegundes, L. e Pereira (2024) sistematizam uma metodologia para o desenvolvimento de estratégias

com *DRL* no mercado futuro, enfatizando validação empírica e controle de risco como etapas essenciais do processo.

Por fim, modelos de *Reinforcement Learning* dialogam com a literatura clássica sobre regimes de mercado, especialmente o arcabouço de Pagan e Sossounov (2003), que estrutura a identificação de períodos de alta e baixa nos mercados financeiros. Assim, a integração entre modelos econométricos de regimes e algoritmos de aprendizado sensíveis a custos consolida uma agenda de pesquisa que combina fundamentos teóricos e técnicas computacionais avançadas para a tomada de decisão em ambientes complexos.

3. Procedimentos Metodológicos

3.1. Modelo de Aprendizado por Reforço

No presente estudo, o problema de otimização de portfólio é expandido ao contexto dos criptoativos. Neste cenário, far-se-á uso da abordagem de *Reinforcement Learning* com uma função de recompensa ajustada ao risco que penaliza a volatilidade para identificar a melhor composição da carteira.

Precisamente, o problema de otimização de portfólio de criptoativos é formulado como um processo de decisão de Markov, no qual o agente de aprendizado por reforço interage com o ambiente (mercado de criptomoedas) em passos de tempo discretos, $t = 1, 2, \ldots, T$. A cada período de tempo, o agente verifica sua ação (ganhos ou perdas) e atualiza seu estado (venda ou compra) do ativo. São definidas as seguintes variáveis no modelo.

Os preços diários de fechamento normalizados dos cinco criptoativos selecionados, *Bitcoin*, *Ethereum*, *Binance Coin*, *Ripple* e *Solana*. O vetor de preço é dado por:

$$s_t = \{P_t^{BTC}, P_t^{ETH}, P_t^{BNB}, P_t^{XRP}, P_t^{SOL}\} \tag{1}$$

O percentual de alocação representa os pesos de alocação do portfólio em cada ativo no instante t. Cada percentual pertence ao espaço contínuo. Precisamente:

$$a_t \in \mathbb{R}^5, \quad a_t \in [min_allocation, max_allocation] \tag{2}$$

A normalização garante que $\sum_i w_{i,t} = 1$ Ademais, permite-se alocação negativa (-40%), correspondendo a posições vendidas.

A função de recompensa é projetada para equilibrar retorno e risco, penalizando a volatilidade excessiva. Formalizando:

$$r_t = \mu_t - \lambda \cdot \sigma_t \tag{3}$$

μ_t é o retorno médio do portfólio; σ_t é o desvio-padrão dos retornos acumulados; e $\lambda = 0,25$ o fator de penalização da volatilidade.

O saldo da carteira, evolução da carteira, é verificado a cada período de tempo. Formalmente:

$$B_{t+1} = B_t \cdot (1 + R_{p,t}) - C_t \tag{4}$$

sendo $R_{p,t} = \sum_i w_{i,t} \cdot r_{i,t}$ o retorno ponderado do portfólio; $r_{i,t}$ o retorno do ativo *i*; e C_t os custos de transação. Os custos são definidos por:

$$C_t = \tau \cdot \|w_t - a_t\|_1 \cdot B_t \tag{5}$$

com $\tau = 0,0005$ (0,05%), o custo proporcional por operação. Assim, os custos são dinâmicos, considerando que este reduz a recompensa quando é realizado em excesso. Por fim, utilizou-se do algoritmo *Proximal Policy Optimization*. Segundo (Schulman *et al.*, 2017), este é adequado para espaços de ação contínuos e ambientes com alta variabilidade.

3.2. Implementação Computacional

Para o melhor funcionamento do modelo, foram definidas algumas condições iniciais e requisitos, a saber: i) Saldo Inicial: O modelo começou com um saldo fictício de $1.000.000 de dólares para simular um portfólio inicial; ii) Criptoativos: Foram selecionadas cinco criptomoedas: *Bitcoin*, *Ethereum*, *Binance Coin*, *Ripple* e *Solana*; iii) Período de Dados: Os dados compreende o período de 1º de janeiro de 2021 a 31 de dezembro de 2024.

Posteriormente, foram definidas algumas condições para tornar o modelo mais fidedigno, cita-se: i) Limites de Alocação: Os Limites de Alocação foram definidos para controlar quanto o gestor pode investir em cada ativo do portfólio, variando entre -40% e 40% do saldo total. Esses limites garantem que o portfólio seja diversificado, evitando que o modelo concentre uma grande parte dos recursos em um único ativo, o que poderia aumentar significativamente o risco. Ademais, os limites permitem que o modelo realize vendas a descoberto (alocações negativas), ou seja, apostar que o preço de um ativo cairá, adicionando flexibilidade à estratégia.

ii) Penalização por Volatilidade: A Penalização por Volatilidade foi incluída para incentivar o modelo a buscar retornos mais estáveis e previsíveis, reduzindo o risco associado a variações bruscas nos preços. A volatilidade, calculada como o desvio padrão dos retornos diários, mede a incerteza dos retornos do portfólio. Um fator de penalização de 0,25 foi aplicado à volatilidade e subtraído da recompensa do modelo, o que significa que estratégias que resultassem em alta volatilidade eram desincentivadas. Essa penalização direcionou o modelo a equilibrar retorno e risco, evitando comportamentos excessivamente agressivos e promovendo uma gestão mais conservadora do portfólio, especialmente em cenários adversos como o *Bear Market*.

iii) Custos de Transação: Foram incluídos custos de 0,05% por transação para simular condições reais desse tipo de mercado.

3.3. Cálculos das Métricas de Desempenho

O modelo foi testado em três cenários distintos: *Bear Market* (mercado em queda), *Sideways Market* (mercado estável) e *Bull Market* (mercado em alta). Para avaliar o desempenho, foram calculadas as seguintes métricas: i) Retorno Total: Definido como a diferença entre o saldo final e o saldo inicial. Essa métrica avalia o ganho ou perda absoluta do portfólio durante o período de teste; ii) Retorno Percentual Acumulado: Busca avaliar a porcentagem de ganho ou perda em relação ao saldo inicial. Este indicador permite uma comparação relativa do desempenho das alocações; iii) Retorno Médio Diário: Média dos retornos diários. Este indicador sinaliza a rentabilidade média do portfólio por dia; iv) Volatilidade Diária: Desvio padrão dos retornos diários. Precisamente, este indicador irá avaliar o risco do portfólio, mostrando a variação dos retornos ao longo do tempo; v) Índice de *Sharpe*: É calculado como a razão entre o retorno médio e a volatilidade diária. Para este indicador, busca-se avaliar o retorno ajustado ao risco, sendo que valores mais altos indicam um melhor desempenho relativo ao risco assumido; vi) Custos de Transação: É calculado o total de custos incorridos durante as transações. Esta métrica indica o impacto das operações de compra e venda no saldo final; vii) *Drawdown*: É obtida para avaliar a máxima perda observada em relação ao pico anterior. Desta maneira, este indicador sinaliza o maior declínio do portfólio durante o período de teste, mostrando o risco de perdas significativas.

Para conferir rigor estatístico à comparação de desempenho das estratégias, aplicou-se o teste não paramétrico de *Kruskal-Wallis* sobre as séries de retornos diários.

A escolha por métodos não paramétricos justifica-se pela não normalidade e pela presença de caudas pesadas características dos retornos de criptoativos. Nos cenários em que a hipótese nula de igualdade distribucional foi rejeitada, utilizou-se o teste de *Mann-Whitney U* para comparações pareadas, adotando-se um nível de significância de 5% ($\alpha = 0,05$).

4. Resultados

4.1. Três cenários de análise

4.1.1. *Bear Market* (5 meses - Janeiro a Junho de 2022)

Os resultados serão apresentados em três cenários de mercado distintos, seguidos de uma análise comparativa final que inclui uma comparação detalhada com a estratégia de alocação estática.

O *Bear Market*, caracterizado por uma tendência de queda prolongada nos preços dos ativos, é um dos cenários mais desafiadores para qualquer estratégia de investimento. O período de janeiro a junho de 2022 (5 meses) foi marcado por um forte declínio no mercado de criptomoedas, impulsionado por uma combinação de fatores macroeconômicos, como o aumento das taxas de juros pelos bancos centrais, a incerteza geopolítica devido à guerra na Ucrânia e o colapso de projetos cripto como o *Terra* (*LUNA*). Esse cenário de desconfiança generalizada e aversão ao risco resultou em quedas significativas nos preços das principais criptomoedas.

As métricas de desempenho para o modelo de *RL* durante o *Bear Market* são:

Tabela 01 – **Métricas de desempenho durante o *Bear Market***

Retorno Total Acumulado	Retorno Médio Diário	Volatilidade Diária	Índice de *Sharpe*	Variância Diária	Custo Total de Transação	Maior *Drawdown* Diário
$-527.867,65	-0,41%	4,27%	-0,0952	0,0018	$97.136,29	58,02%

Fonte: Elaboração Própria.

O desempenho individual de cada uma das cinco criptomoedas analisadas durante o *Bear Market* é: *BTC*: -33,33%; *ETH*: -48,48%; *BNB*: -39,23%; *SOL*: -74,36%; *XRP*: -50,34%. Esses valores representam a variação percentual no preço de cada ativo no período.

Ao visualizar a tabela 1, verifica-se que no *Bear Market*, o modelo de *RL* apresentou um retorno acumulado de -52,79%, desempenho ligeiramente inferior ao da alocação estática de 20% em cada criptomoeda, que teve um retorno de -50,21%. Ambos os resultados foram negativos, como esperado em um mercado em queda, mas o modelo de *RL* não conseguiu superar a estratégia de alocação estática. Em comparação com o desempenho individual do *Bitcoin* (*BTC*), que teve uma queda de -33,33%, tanto o modelo de *RL* quanto a alocação estática tiveram desempenho pior. Estes resultados sugerem que em um cenário de *Bear Market*, concentrar-se apenas no *Bitcoin* é uma estratégia mais eficaz do que diversificar o portfólio ou utilizar o modelo de *RL*.

O Índice de *Sharpe* do modelo de *RL* foi -0,0952, indicando que o modelo não é eficiente em gerar retornos ajustados ao risco durante esse período. Além disso, o alto custo de transação $ 97.136,29, sugere que o modelo realizou um número significativo de operações, possivelmente tentando ajustar a alocação do portfólio para reduzir as perdas, mas sem sucesso. Em suma, no *Bear Market*, o modelo de *RL* não conseguiu superar nem a alocação estática nem o desempenho individual do *Bitcoin*.

Tabela 02 – **Desempenho de cada estratégia durante o *Bear Market***

Modelo de *RL*	Somente *BTC*	Alocação 20%
-52,79%	-33,33%	-50,21%

Fonte: Elaboração Própria.

4.1.2. *Sideways Market* (2 Meses - Setembro a Novembro de 2002)

O *Sideways Market*, ou mercado lateral, é caracterizado por uma tendência de preços estáveis, sem movimentos significativos de alta ou baixa. O período de setembro a novembro de 2022 (2 meses) é marcado por uma relativa estabilidade no mercado de criptomoedas, após os fortes movimentos de baixa observados no primeiro semestre do ano. Durante esse período, os investidores adotaram uma postura mais cautelosa, aguardando sinais claros de recuperação ou novas quedas. A ausência de grandes catalisadores de mercado resultou em uma fase de consolidação, com os preços das criptomoedas oscilando dentro de uma faixa estreita.

As métricas de desempenho para o modelo de *RL* durante o *Sideways Market* são:

Tabela 03 – **Métricas de desempenho durante o *Sideways Market***

Retorno Total Acumulado	Retorno Médio Diário	Volatilidade Diária	Índice de *Sharpe*	Variância Diária	Custo Total de Transação	Maior *Drawdown* Diário
$ 36.381,49	0,096%	2,70%	0,0357	0,00073	$ 47.401,00	16,86%

Fonte: Elaboração Própria.

O desempenho individual de cada uma das cinco criptomoedas analisadas durante o *Sideways Market* são: *BTC*: 1,83%; *ETH*: -0,85%; *BNB*: 17,22%; *SOL*: 3,25%; *XRP*: 39,83%. A alocação estática durante o *Sideways Market* apresentou um retorno de 12,55%.

No *Sideways Market*, o modelo de *RL* apresentou um retorno acumulado de 3,64%, desempenho inferior ao da alocação estática de 20% em cada criptomoeda, que teve um retorno de 12,55%. O modelo de *RL* também teve um desempenho superior ao do *Bitcoin* (*BTC*), que teve um retorno de 1,83%, mas não conseguiu capitalizar totalmente o desempenho excepcional de *XRP* (39,83%), que foi o principal impulsionador do retorno da alocação estática.

O Índice de *Sharpe* do modelo de *RL* foi 0,0357, indicando que o modelo gerou retornos ajustados ao risco, porém de forma modesta. O custo de transação $ 47.401,00 é menor que no *Bear Market*, sugerindo uma menor frequência de operações, o que é esperado em um mercado lateral. Em resumo, no *Sideways Market*, o modelo de *RL* teve um desempenho melhor do que o *Bitcoin*, mas não conseguiu superar a alocação estática, que se beneficiou do forte desempenho de *XRP*.

Tabela 04 – **Desempenho de cada Estratégia durante o *Sideways Market***

Modelo de *RL*	Somente *BTC*	Alocação 20%
3,64%	1,83%	12,55%

Fonte: Elaboração Própria.

4.1.3. *Bull Market* (6 Meses - Julho de 2024 a Janeiro de 2025)

O *Bull Market*, caracterizado por uma tendência de alta prolongada nos preços dos ativos, é o cenário mais favorável para estratégias de investimento. O período de julho de 2024 a janeiro de 2025 (6 meses) foi simulado como um cenário de alta, no qual o mercado de criptomoedas experimentou uma recuperação significativa, impulsionada por fatores como a adoção institucional crescente, a regulamentação favorável e o aumento da confiança dos investidores. Durante esse período, os preços

das criptomoedas apresentaram tendências de alta, com alguns ativos, como *XRP*, registrando ganhos expressivos.

As métricas de desempenho para o modelo de *RL* durante o *Bull Market* são:

Tabela 05 – **Métricas de desempenho durante o *Bull Market***

Retorno Total Acumulado	Retorno Médio Diário	Volatilidade Diária	Índice de *Sharpe*	Variância Diária	Custo Total de Transação	Maior *Drawdown* Diário
$ 732.641,81	0,35%	3,13%	0,1117	0,00098	$ 179.779,98	35,96%

Fonte: Elaboração Própria.

O desempenho individual de cada uma das cinco criptomoedas analisadas durante o *Bull Market* é: *BTC*: 48,65%; *ETH*: -3,13%; *BNB*: 21,55%; *SOL*: 29,11%; *XRP*: 336,41%. O Retorno Percentual Final da Carteira é de 66,62%.

Ao verificar a tabela 5, nota-se que no *Bull Market*, o modelo de *RL* apresentou um retorno acumulado de 73,26%, superando tanto o desempenho do *Bitcoin* (*BTC*), que teve um retorno de 48,65%, quanto o da alocação estática de 20% em cada criptomoeda, que teve um retorno de 66,62%. Desta forma, o modelo de *RL* foi capaz de capitalizar o forte desempenho de *XRP* (336,41%), que foi o principal impulsionador dos retornos, enquanto também mitigou as perdas de *Ethereum* (-3,13%).

O Índice de *Sharpe* do modelo de *RL* foi 0,1117, indicando que o modelo é eficiente em gerar retornos ajustados ao risco durante esse período. No entanto, o custo de transação $ 179.779,98 é o mais alto entre os três cenários, refletindo a maior frequência de operações realizadas pelo modelo para capitalizar as tendências de alta. Em resumo, no *Bull Market*, o modelo de *RL* superou tanto o *Bitcoin* quanto a alocação estática, demonstrando sua eficácia em mercados de alta.

Tabela 06 – **Desempenho de cada Estratégia durante o *Bull Market***

Modelo de *RL*	Somente *BTC*	Alocação 20%
73,26%	48,65%	66,62%

Fonte: Elaboração Própria.

4.2. Análise Comparativa Entre os Cenários

Uma análise comparativa final entre os três cenários de mercado (*Bear*, *Sideways* e *Bull*) permite avaliar a eficácia do modelo de *Reinforcement Learning* em diferentes condições, comparando seu desempenho com duas estratégias tradicionais: investir apenas em *Bitcoin* e adotar uma alocação estática de 20% em cada uma das cinco criptomoedas analisadas. Essa comparação é essencial para entender em quais cenários o modelo de *RL* se destaca e onde ele pode precisar de ajustes.

Os resultados evidenciam que a eficácia do modelo de *Reinforcement Learning* (*RL*) depende fortemente do regime de mercado analisado. Essa constatação é coerente com a literatura, que aponta que modelos de *RL* tendem a sobressair em mercados de tendência clara (*bull markets*), mas apresentam limitações em cenários de queda ou lateralização quando não incorporam mecanismos explícitos de gestão de risco em sua função de recompensa (Sood *et al*., 2023; Huang *et al*., 2024).

No *Bear Market*, o modelo de *RL* apresentou um retorno acumulado de -52,79%, desempenho ligeiramente pior do que a alocação estática de 20% em cada criptomoeda, que teve um retorno de -50,21%. Ambos os resultados foram negativos, como esperado em um mercado em queda, mas o modelo de *RL* não conseguiu superar a estratégia de alocação estática. Em comparação com o desempenho individual do *Bitcoin*, que teve uma queda de -33,33%, tanto o modelo de *RL* quanto a alocação estática tiveram desempenho pior. Esses resultados sugerem que, em um cenário de

Bear Market, concentrar-se apenas no *Bitcoin* poderia ter sido uma estratégia mais eficaz do que diversificar o portfólio ou utilizar o modelo de *RL*.

O período de Janeiro a Junho de 2022 foi marcado por uma forte desconfiança no mercado de criptomoedas, com eventos como o colapso do *Terra* (*LUNA*) e o aumento das taxas de juros globais, que levaram a uma aversão ao risco generalizada. O modelo de *RL* teve um custo de transação elevado ($ 97.136,29), o que pode ter contribuído para o desempenho inferior em relação ao *Bitcoin* e à alocação estática. Além disso, o modelo tentou ajustar a alocação do portfólio para reduzir as perdas, mas não foi eficiente o suficiente para superar a queda generalizada dos preços. A dificuldade em mitigar perdas em um cenário de forte declínio sugere que o modelo pode precisar de ajustes para melhorar sua eficácia em mercados de baixa.

O resultado encontrado dialoga com Cui *et al.* (2023), que mostram que abordagens de *RL* sem ajustes de risco explícitos (como o *CVaR*) tendem a falhar na contenção de perdas extremas, dada a presença de caudas pesadas no mercado de criptoativos. O elevado custo de transações registrado também está em linha com os encontrados no estudo de Mishra, Srivastava e Singh (2025), que destacam que estratégias de *RL* sem penalizações adequadas podem incorrer em sobre ajuste e excesso de rebalanceamentos, comprometendo a eficiência líquida.

No *Sideways Market*, o modelo de *RL* apresentou um retorno acumulado de 3,64%, desempenho inferior ao da alocação estática de 20% em cada criptomoeda, que teve um retorno de 12,55%. No entanto, o modelo de *RL* superou o desempenho do *Bitcoin*, que teve um retorno de 1,83%. A alocação estática se beneficiou do forte desempenho de *XRP* (39,83%), enquanto o modelo de *RL* não conseguiu capitalizar totalmente esse desempenho excepcional.

O período de Setembro a Novembro de 2022 foi marcado por uma relativa estabilidade no mercado, com os investidores adotando uma postura mais cautelosa após os fortes movimentos de baixa do primeiro semestre de 2022. O retorno elevado da *XRP* foi o principal impulsionador do desempenho da alocação estática, enquanto o modelo de *RL* não conseguiu alocar uma parcela significativa do portfólio nesse ativo. O custo de transação do modelo de *RL* $ 47.401,00 foi menor que no *Bear Market*, mas ainda pode ter impactado negativamente o retorno final. Apesar de o modelo ter superado o *Bitcoin*, sua incapacidade de capitalizar o desempenho de *XRP* sugere que ele pode precisar de ajustes para melhorar a alocação em ativos de alto desempenho em mercados laterais.

Esse comportamento reforça o argumento de que modelos de *Reinforcement Learning* focados apenas em retorno ajustado por *Sharpe* podem perder oportunidades específicas de ativos individuais (Huang *et al*., 2024). Por outro lado, a redução nos custos de transação nesse período indica que o modelo foi mais cauteloso no rebalanceamento, ainda que sem capturar os ganhos plenos do mercado lateral.

Por fim, no *Bull Market*, o modelo de *Reinforcement Learning* apresentou um retorno acumulado de 73,26%, superando tanto o desempenho do *Bitcoin* (48,65%) quanto o da alocação estática de 20% em cada criptomoeda, que teve um retorno de 66,62%. O modelo de *RL* foi capaz de capitalizar o forte desempenho de *XRP* (336,41%), que foi o principal impulsionador dos retornos, enquanto também mitigou as perdas de *Ethereum* (-3,13%).

O período de Julho de 2024 a Janeiro de 2025 foi simulado como um cenário de alta, com o mercado de criptomoedas experimentando uma recuperação significativa, impulsionada por fatores como a adoção institucional crescente e a regulamentação favorável. O modelo de *RL* demonstrou ser eficiente em identificar e alocar recursos em ativos com desempenho excepcional, como *XRP*, enquanto reduzia a exposição a ativos com desempenho negativo, como *Ethereum*. Apesar de o custo de transação ter sido

o mais alto entre os três cenários $ 179.779,98, o modelo de *RL* ainda conseguiu gerar retornos superiores, indicando que a estratégia de rebalanceamento dinâmico foi eficaz em um mercado de alta. A capacidade do modelo de capitalizar ganhos em um cenário de alta foi crucial para seu sucesso, demonstrando sua eficácia em mercados favoráveis.

Os resultados obtidos no estudo estão de acordo com Sood *et al.* (2023), que evidenciam a vantagem do *Reinforcement Learning* em mercados de tendência positiva, dado que agentes de política adaptativa conseguem identificar e explorar ativos com retornos excepcionais. Além disso, o modelo foi capaz de reduzir a exposição a ativos com desempenho negativo, como *Ethereum*, confirmando a observação de Liu *et al.* (2023) de que *Reinforcement Learning* se destaca ao ajustar dinamicamente as alocações em mercados de crescimento.

A descoberta mais contundente desta pesquisa reside na análise de significância estatística longitudinal. Independentemente da conjuntura macroeconômica e do viés direcional do mercado, o teste não paramétrico de *Kruskal-Wallis* indicou p-valores consistentemente superiores ao nível de significância de 5% (0,9696 no *Bear*, 0,9209 no *Sideways* e 0,7437 no *Bull*). Consequentemente, a hipótese nula (H_o) foi aceita em 100% dos regimes testados.

5. Considerações Finais

Buscando verificar se a abordagem de *Reinforcement Learning* possibilita melhor desempenho que estratégias estáticas de investimento, o presente estudo encontrou que o modelo de *RL* apresentou um desempenho variável, dependendo das condições de mercado. Precisamente, no *Bear Market*, o modelo não foi eficaz em mitigar perdas, superando apenas marginalmente o *Bitcoin* e ficando aquém da alocação estática. Esse resultado sugere que, em cenários de baixa, estratégias mais conservadoras, como a concentração em ativos consolidados como o *Bitcoin*, podem ser mais adequadas.

No *Sideways Market*, o modelo de *RL* superou o *Bitcoin*, mas não conseguiu superar a alocação estática, que se beneficiou do forte desempenho de *XRP*. Os resultados obtidos indicam que, em mercados laterais, a diversificação pode ser mais vantajosa do que a concentração em um único ativo, embora o modelo de *RL* ainda precise de aprimoramentos para capitalizar de forma mais eficiente os ativos com desempenho excepcional. No *Bull Market*, o modelo de *RL* demonstrou sua eficácia, superando tanto o *Bitcoin* quanto a alocação estática. A capacidade do modelo de identificar e alocar recursos em ativos com forte desempenho foi determinante para o sucesso nesse cenário.

Os resultados obtidos reforçam a importância de adaptar estratégias de investimento às diferentes condições de mercado. O modelo de *RL* mostrou-se mais eficiente em mercados de alta, no qual sua capacidade de identificar tendências e alocar recursos em ativos promissores foi decisiva. Contudo, em mercados de baixa e laterais, o modelo enfrentou desafios para superar estratégias mais conservadoras, como a concentração em *Bitcoin* ou a alocação estática. Estes resultados sugerem que o modelo pode se beneficiar através de alguns fatores, como a redução de custos de transação, a incorporação de mecanismos de proteção contra quedas bruscas e a otimização da alocação em ativos de alto desempenho em cenários adversos.

Além disso, fatores externos, como a conjuntura macroeconômica e eventos específicos do mercado de criptomoedas (por exemplo, o colapso do *Terra/LUNA*), tiveram um impacto significativo nos resultados, destacando a necessidade de incorporar variáveis macroeconômicas e análises de risco em futuras versões do modelo. A inclusão de técnicas de aprendizado por reforço profundo (*Deep Reinforcement Learning*) e a exploração de modelos híbridos, que combinem *RL*

com outras abordagens de *machine learning* podem ser caminhos promissores para melhorar a robustez do modelo em diferentes condições de mercado.

Este estudo contribui para a literatura ao demonstrar como técnicas de *Reinforcement Learning* podem ser aplicadas na otimização de portfólios de criptoativos em diferentes cenários de mercado (*Bull*, *Bear* e *Sideways*), algo ainda pouco explorado em pesquisas anteriores. Ao comparar o desempenho do modelo de *RL* com estratégias tradicionais, como alocação estática e concentração em *Bitcoin*, evidencia-se que o *RL* é particularmente eficaz em mercados de alta, mas enfrenta limitações em cenários de baixa e laterais, oferecendo *insights* sobre os fatores que podem melhorar sua robustez, como redução de custos de transação, mecanismos de proteção contra quedas bruscas e alocação dinâmica em ativos de alto desempenho.

Ademais, a abordagem do estudo amplia a literatura metodológica ao integrar retorno ajustado ao risco com *Reinforcement Learning*, destacando caminhos para futuras pesquisas que combinem *RL* com modelos híbridos, dados macroeconômicos e estratégias de diversificação dinâmica, reforçando a aplicabilidade do *RL* na gestão de portfólios em mercados voláteis e imprevisíveis.

Por fim, sugere-se a realização de estudos adicionais para avaliar o desempenho do modelo em períodos mais extensos e em diferentes classes de ativos, bem como a investigação de estratégias de diversificação dinâmica que possam adaptar-se rapidamente às mudanças de mercado. A integração de dados em tempo real e a consideração de fatores comportamentais e sentimentais do mercado também podem contribuir para o aprimoramento do modelo, tornando-o mais eficaz em cenários complexos e voláteis. Essas melhorias têm o potencial de ampliar a aplicabilidade do modelo de *RL* no gerenciamento de portfólios e na tomada de decisões de investimento.

O modelo pode ser acessado em: https://github.com/vaufer/Reinforcement-Learning-with-Statistical-Analysis

Referências

BLACK, F.; LITTERMAN, R. Global Portfolio Optimization, p. 28-43. 1992.

BÖHME, R. *et al.* Bitcoin: Economics, Technology, and Governance, 2007.

BUTERIN, V. Ethereum Whitepaper. 2013.

COINMARKETCAP. Cryptocurrency Prices by Market Cap. Disponível em: https://coinmarketcap.com/. Acesso em: Julho, 2025.

Conegundes, L.; Pereira, A. C. A Methodology for Developing Deep Reinforcement Learning Trading Strategies: A Case Study in the Futures Market. In 2024 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics, 2024.

CUI, T.; DING, S.; JIN, H.; ZHANG, Y. Portfolio constructions in cryptocurrency market: A CVaR-based deep reinforcement learning approach. Economic Modelling, v. 119, 2023.

DOWLING, M. Is non-fungible token pricing driven by cryptocurrencies?. 2022. EL SALVADOR GOVERNMENT. Ley Bitcoin. 2021. Disponível em: https://www.presidencia.gob.sv/ley-bitcoin-y-su-entorno-de-implementacion-aceleran-la-transformacion-digital-de-el-salvador/. Acesso em: Julho, 2025.

FAMA, E.; FRENCH, K. The Cross-Section of Expected Stock Returns. The Journal of Finance, p. 427-46, 1992.

FEDERAL RESERVE (FED). Monetary Policy Report. 2022.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep Learning. 2017.

Gu, Y.; Cheng, F.; Yang, L.; Xu, J.; Chen, X.; Cheng, L. Cost-aware cloud workflow scheduling using DRL and simulated annealing. Digital Communications and Networks, 10(6), 1590-1599, 2024.

Liu, P. Efficient Portfolio Rebalancing via Cost-aware Bayesian Optimization. 2024.

LIU, X.; ZHANG, W.; CHEN, X. A survey of deep learning applications in cryptocurrency. Frontiers in Physics, 2023 (PMC).

Lu, J. Y.; Lai, H. C.; Shih, W. Y.; Chen, Y. F.; Huang, S. H.; Chang, H. H.; Dai, T. S. Structural break-aware pairs trading strategy using deep reinforcement learning. The Journal of Supercomputing, 78(3), 2022.

MARKOWITZ, H. Portfolio Selection: Efficient Diversification of Investments. Yale University Press, f. 185, 1952.

MISHRA, A.; SRIVASTAVA, A.; SINGH, S. Risk-Adjusted Deep Reinforcement Learning for Portfolio Optimization. Computational Economics, 2025.

NAKAMOTO, S. Bitcoin, a Peer-to-peer Electronic Cash System. 2008.

Pagan, A. R.; Sossounov, K. A. A simple framework for analysing bull and bear markets. Journal of applied econometrics, 18(1), 23-46, 2003.

SCHÄR, F. Decentralized Finance: On Blockchain- and Smart Contract-Based Financial Markets, p. 153-174, 2021.

SHARPE, W. F. Mutual fund performance. The Journal of Business, v. 39, n. 1, p. 119-138, 1966.

SHARPE, W. F. The Sharpe ratio. The Journal of Portfolio Management, v. 21, n. 1, p. 49-58, 1994.

SOOD, S.; PAPASOTIRIOU, K.; VAICIULIS, M.; BALCH, T. Deep Reinforcement Learning for Optimal Portfolio Allocation: A Comparative Study with Mean-Variance Optimization. In: ICAPS 2023 – FinPlan Workshop. 2023.

SUTTON, R. S.; BARTO, A. G. Reinforcement learning: an introduction. 2. ed. Cambridge: MIT Press, 2018.

SUTTON, R. S.; BARTO, A. G. Reinforcement Learning: An Introduction. MIT Press, 1997.

HUANG, Y. *et al.* A Self-Rewarding Mechanism in Deep Reinforcement Learning for Trading Strategy Optimization. Mathematics, v. 12, n. 24, p. 4020, 2024.

ZHENG, Z. An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends. 2017.